Annals of R.S.C.B., ISSN:1583-6258, Vol. 24, Issue 2, 2020, Pages. 270 - 277
Received 24 October 2020; Accepted 15 December 2020.

Air Quality Index Forecasting for Assessing the Status of Air Quality Using Ann & Lstm Neural network

Prof. Mona Mulchandani[1], Ms. Priyanka Dudhe[2], Ms. Sakshi Jaswani[3]
[1][2]Assistant Professor, Jhulelal Institute of Technology, Nagpur, India
Nagpur, India
[3]Student, Jhulelal Institute of Technology, Nagpur, India
Nagpur, India

ABSTRACT

As the impact of air pollutants is increasing drastically on living beings in recent years, it has become very necessary to address the issue of air pollution control by scientists and environmental lists. To ensure the same, it is important to forecast air quality in terms of parameters that cause air pollution directly or indirectly, which generally affects the living population. Environmental Protection Agency (USEPA) has suggested a method to estimate air quality index in a region which constitutes different concentration of criteria air pollutants such as RSPM, SO2, NO2, and SPM. The motive of our research is to model a predicting model for forecasting daily AQI that can be put to use for local and regional air quality management.

Keywords: AQI, AR, ARIMA**.**

INTRODUCTION

It is critical to do city air quality forecasting work in order to reduce air pollution in urban areas and to enhance the living environment of city people. The air quality index (AQI) is a one-dimensional metric that quantifies the status of air quality. Air pollution is becoming a greater hazard to civilization, and different measures have lately been implemented to combat it. When the weather is overcast or hot tube, the air Quality index will forecast it. The issue of accurate air pollution forecast remains a source of worry. Because the AQI has a predictable pattern, deep learning models may be utilized to accurately forecast future AQI values.

The forget gate's goal is to delete past data that isn't relevant to the future forecast. The BP neural network, also known as an artificial neural network, is the algorithm that will be utilized. The traditional back-propagation approach is used to train this network. This paper's objective is to sample time-series data at discrete time points with a consistent time interval. BP neural network and genetic algorithm were utilized as algorithms. Website data may be collected in a format using web scraping technology. It can collect data using machine learning methods. The procedure of data cleaning and preparation is then applied.

As a result, the Air Quality Index technique will be more accurate. The forget gate's goal is to delete past data that isn't relevant to the future forecast. The BP neural network and the genetic algorithm will be utilized as algorithms. The traditional back-propagation approach is used to train this network[1].

Annals of R.S.C.B., ISSN:1583-6258, Vol. 24, Issue 2, 2020, Pages. 270 - 277
Received 24 October 2020; Accepted 15 December 2020.

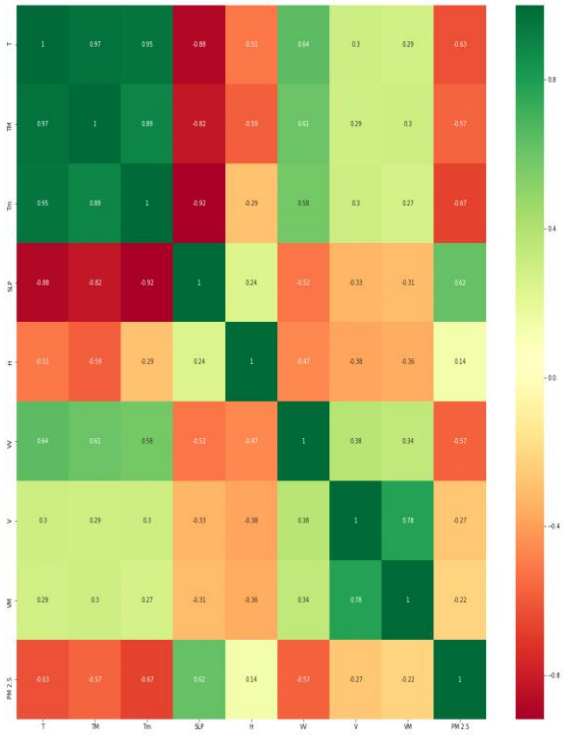

Fig1-Artificial Neural Network

RELATED WORK

This section provides a literature review for air quality research and its status, as well as the methods to conduct air quality research. Distal used the two-pollutant Cox proportional hazards model to assess mortality associated with exposure to PM2.5 and ozone.The findings of the study show that PM2.5 and ozone exposure might have negative consequences, which are more evident among ethnic minorities and low-income persons. Raaschou-Nielsenetal. 4 estimated the association between the components of a "particular matter" and lung cancer incidence. They suggested that the effect on lung cancer depends on the composition of a particular matter. For example, inhalation of contaminating particles containing Ni and S elements can have a more adverse impact. Inhaled PM may induce adverse cardiovascular reactions through three potential mediators. It is impossible to avoid air pollution altogether, but the current restrictive standards should be guaranteed to reduce the source of potentially polluting airparticles. [2]

Annals of R.S.C.B., ISSN:1583-6258, Vol. 24, Issue 2, 2020, Pages. 270 - 277
Received 24 October 2020; Accepted 15 December 2020.

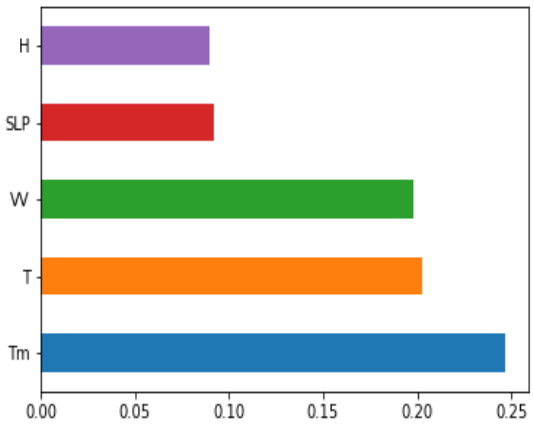

Fig2 -K Nearest Neighbor Regression

Inhaled. It has been proposed that particle air pollutants have a substantial link with stroke mortality, and that environmental health measures that reduce air pollution can lower stroke risk. In terms of air quality research techniques, AI-based algorithms combined with big data mining and analysis may unearth useful data and uncover associated correlations and decision-

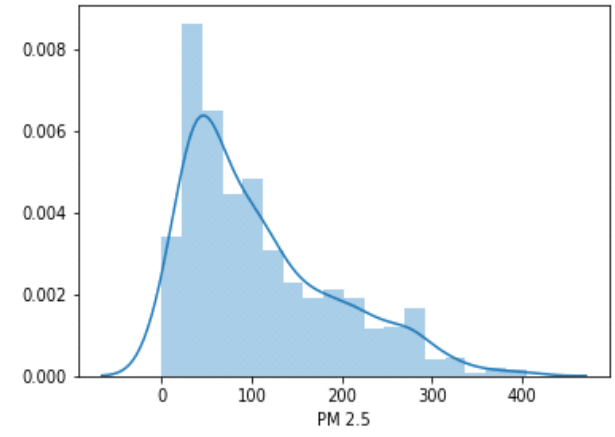

Fig3-Linear Regression

The most frequent type of study is excavation around a single city. Furthermore, during the previous two decades, the utilization of topographical and meteorological data has evolved. Furthermore, socioeconomic elements such as gross domestic product (GDP) and rapid population growth are gradually incorporated in the mechanism analysis. In China, there are 87 cities with the widest variety of research. Kalapanidas is a NEMO prototype that was created to enable short-term nitrogen dioxide forecasting. Using a lazy learning technique, the case-based reasoning system, they categorised pollution levels into four levels (a) low, (b) medium, (c) high, and (d) alert.

The air quality index prediction model is built using the enhanced BP neural network, and the AQI value of the day can be predicted using the weather forecast for the day and the AQI value from the day before yesterday. The single BP neural network method is prone to falling into the local minimum of the error function, therefore the evolutionary algorithm is employed to

Annals of R.S.C.B., ISSN:1583-6258, Vol. 24, Issue 2, 2020, Pages. 270 - 277
Received 24 October 2020; Accepted 15 December 2020.

develop the BP neural network, which can compensate for its flaws. The enhanced BP neural network predicts the air quality index with an accuracy rate of 80.44 percent, and the air quality level with an accuracy rate of 82.5 percent, according to the experimental study. The prediction results of the BP neural network are superior than those of a single neural network.

The model has a particular reference value, and it may be used to avoid and regulate urban air pollution. Artificial neural networks, commonly referred to as "neural networks," are computing systems that are based on the biological neural networks that make up animal brains. An ANN is made up of artificial neurons, which are a collection of linked units or nodes that loosely resemble the neurons in a biological brain[3].

METHODOLOGY

ANN- Artificial Neural Network

Artificial neural networks, or simply neural networks, are computing systems that are loosely based on the biological neural networks that make up animal brains. Artificial neurons are a set of linked units or nodes in an ANN that loosely resemble the neurons in a biological brain. [3].

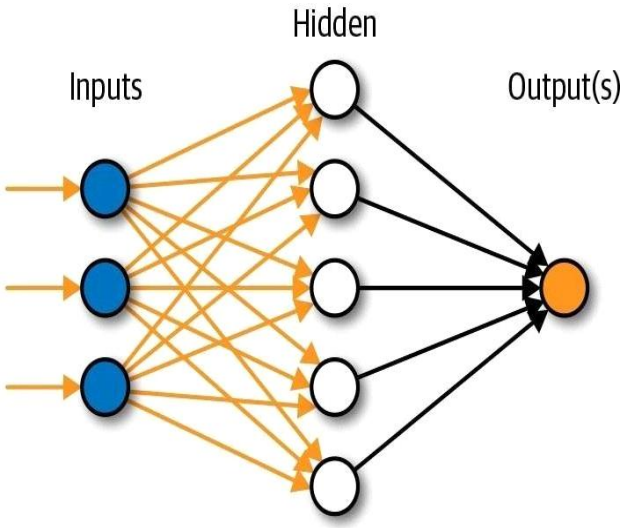

Fig 4– Artificial Neutral Network

Features Extraction

Types Preliminaries of ANN

An artificial neural network (ANN) can be thought of as a linked collection of basic processing components (or units/nodes/neurons). The inter-unit connection strengths or weights generated through a process of learning from a series of training patterns record the network's processing capabilities. One input layer, one output layer, and hidden layers make up a standard ANN.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 24, Issue 2, 2020, Pages. 270 - 277
Received 24 October 2020; Accepted 15 December 2020.

Each layer can have many units, each of which has an output that is a function of the weighted sum of its inputs. The weighted total of outputs from nodes connected to it is the input into a node.

Each unit uses an activation function to process its net input. However, Cybenko (1989) demonstrated that an ANN can represent every continuous function (provided there are enough hidden units) with one hidden layer, and any function with two hidden layers. Similar to coefficients in a regression model, the weights in an ANN are adjusted to solve the of these weights.

There are two forms of ANN learning: supervised and unsupervised learning. When each input in the training set has a known goal value, it is referred to as supervised learning. The ANN's output is compared to a target value, and the difference is utilized to train the algorithm (alter the weights). When there are no desired output values corresponding to input patterns in the training data, unsupervised learning is required. [4].

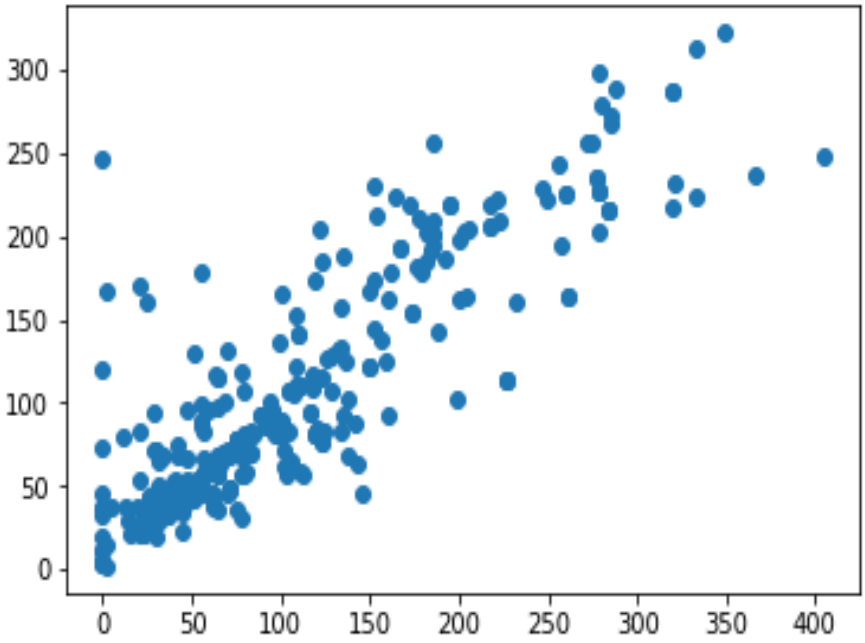

Fig5-Random Forest Regressor

Multilayered feed forward Artificial Neural Network (MLFANN)

Units in one layer are only connected to units in the next layer, not to units in the previous layer or units in the same layer, in an MLFANN. An MLFANN can have several hidden layers, each with a different number of hidden units. It is standard practice not to consider the input layer when counting layers because it does not do any calculation and just transmits data to the next layer. A two-layered MLFANN is one that has an input layer, one hidden layer, and one output layer.

The MLFANN network design is the most often used. It's a layered feed forward network with units organized in a layered feed forward architecture. As a result, the network may be

Annals of R.S.C.B., ISSN:1583-6258, Vol. 24, Issue 2, 2020, Pages. 270 - 277
Received 24 October 2020; Accepted 15 December 2020.

interpreted simply as an input-output model with weights and thresholds (biases) as free parameters. The number of layers and units in each layer determine the function complexity, allowing such networks to represent functions of practically any complexity. [5].

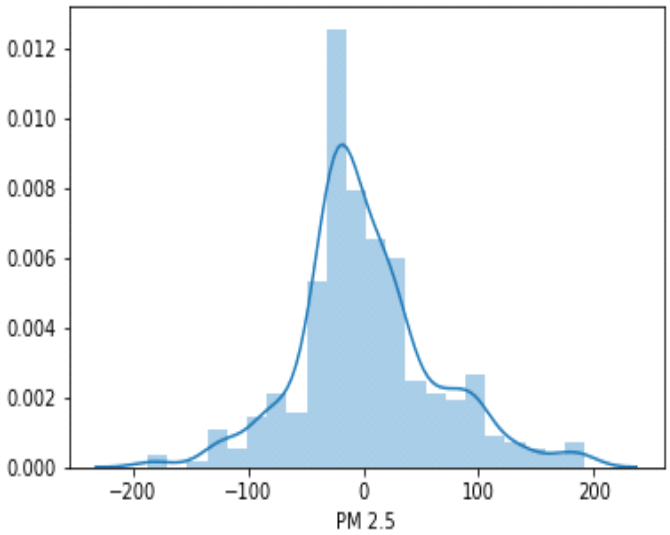

Fig6-Ridge and Lasso Regression

Learning Algorithms

A learning method modifies connection weights when the input–output vectors are given to the network, until the system converges on a function that properly reproduces the output. With the goal of reducing the sum of the squared error functions of the network output, the gradient descent algorithm or conjugate gradient descent method may be used to find optimal connection weights. [5][6].

Datasets Description

The entry of data sets is the initial stage. To further test the efficiency of our technique, three data sets are utilized for feature selection: data collection, data processed, and data clean, all of which are available from Web Scrapping Technology.

Data collection is the process of obtaining and analyzing data on certain variables in a structured manner, allowing one to answer pertinent questions and assess consequences[5][6].

Lasso regression is a kind of shrinkage-based linear regression. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regressionis well- suited for models showing high levels of multi collinearity or when you want to automate certain parts of model selection, like variable selection/parameterelimination[7].

A mathematical model having a number of parameters that must be learnt from data is referred to as a Machine Learning model. We may fit the model parameters by training a model using

Annals of R.S.C.B., ISSN:1583-6258, Vol. 24, Issue 2, 2020, Pages. 270 - 277
Received 24 October 2020; Accepted 15 December 2020.

existing data. [6] Hyper parameters, on the other hand, are a type of parameter that cannot be learned directly from the usual training procedure. They are generally addressed prior to the start of the training procedure.

Data Sets	No of classes	No of Instance	No of Features
AQI_2013	9	370	8
AQI_2014	9	420	5
AQI_2015	9	350	7

Table 1- Dataset Description For Dataset Used In EGFS

The ability of artificial neural network approach to properly address the genuine case when the precise nonlinear functional connection between the response variable and a collection of predictors is unknown has been highlighted. Although ANNs may not be capable of providing the same amount of insight as many statistical models, treating them as "black boxes" is incorrect.In fact, one of the most active study areas in ANN is "understanding the influence of predictors on response variable." It is envisaged that in the future, researchers would use not just MLFANN but also more sophisticated ANN models such as "Radial basis function neural network" and "Generalized regression neural network" in their study [5][7].

CONCLUSION

In this paper, we have built, examined and carried out a comparative study of a hybrid neural network model with LSTM to forecast Air Quality Index. The results of our model have been compared to standalone neural network models to show the enhancement of forecast. Some variations of the hybrid model have also been suggested to improve the overall performance of the model and reduce the computation time of the forecast. The model has been tested with data gathered from a single station however, the same procedure can be extended to forecast AQI of other cities and regions as well. As a part of further development in this work, the model is desired to be included in an application in mobile devices for use in everyday lives.

REFERENCES

1. “Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in Lanzhou“,2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), DOI: 10.1109/CIAPP.2017.8167199.

2. “Air Quality Index forecasting using ANN and LSTM Neural Network”, Vinod KundlikGajbhiye,Prof. JayantAdhikari, International Journal of Innovations in Engineering and Science, Vol. 5, No. 12, 2020 , PP. 20-23

3. http://ijies.net/finial-docs/finial-manuscripts/21314311012021612.docx

4. http://ijies.net/finial-docs/finial-manuscripts/141533110120215125.docx

5. “Application ANN: A Review” , Anjali, International Journal of Engineering Research & Technology (IJERT) IJERT www.ijert.org NCAEM-2013 Conference Proceedings ISBN: 978-93-83758-09-8

6. S. N. Ajani and S. Y. Amdani, "Probabilistic path planning using current obstacle position in static environment," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170727.

7. D. Veeraiah and J. N. Rao, "An Efficient Data Duplication System based on Hadoop Distributed File System," 2020 International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 197-200, doi: 10.1109/ICICT48043.2020.9112567.

8. Rao, J. Nageswara, and M. Ramesh. "A Review on Data Mining & Big Data." *Machine Learning Techniques. Int. J. Recent Technol. Eng* 7 (2019): 914-916.

9. Karthik, A., MazherIqbal, J.L. Efficient Speech Enhancement Using Recurrent Convolution Encoder and Decoder. Wireless Pers Commun (2021). https://doi.org/10.1007/s11277-021-08313-6

10. S. Ajani and M. Wanjari, "An Efficient Approach for Clustering Uncertain Data Mining Based on Hash Indexing and Voronoi Clustering," 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 486-490, doi: 10.1109/CICN.2013.106.